NOAA Institutional Repository – TEAM on Policy, Metadata, etc.
Draft Recommendations as of April 3, 2006

Team Members:
Anna Fiolek - leading
Doria Grimes
Stanley Elswick
Dottie Anderson
Erie Taniuchi
Janine Devereaux – on call
Don Collins
Linda Preston – on call
Gulnar Nagashybayeva
Margo Menconi

Policy issues:

1. **CONTENT and content submission criteria.**
   All NOAA documents currently online and in the public domain should be considered for inclusion in the NOAA Institutional Repository (NOAA IR) database. The priority should be given to the "official" NOAA document being in the public domain, cleared of copyrights (i.e., the collaborative work documents). These documents are:
   - Official NOAA technical report series (tech. memos, data reports, staff reports, etc.) currently available online.
   - NOAA employees' peer-reviewed journal articles, pre-prints, research papers, conference papers, gray literature, etc.
   - NOAA subjectively selected websites, home pages, digital collections of historical importance (all formats, videos, images, audio recordings, maps, etc.; consider to include scientific datasets info?)
   - New items NOAA offices and NOAA employees publish officially online, including annual reports, procedures, guidelines and regulatory papers.
   - New and historic NOAA publications selected for digitization by NOAA IR team or other NOAA personnel.
   - NOAA publications currently available online through GPO or NTIS (NOAA electronic fugitive documents)

Content should be both cumulative and maintain its long-term access. Therefore, the submission criteria should also be considered. Items once submitted cannot be withdrawn, except in presumably rare cases (See below section 6. Withdrawals).

The priority aspect should be considered. Documents deemed to be the most valuable, up-to-date, and/or helpful-to-NOAA-users should be designated as the highest priority for inclusion.

Also, NOAA IR shall aim in preserving and making accessible its digital content on a long-term basis. Digital preservation and long-term access are tightly linked together, each being meaningless without the other. NOAA IR should dictate the content submission standard

(.pdf, .txt, .doc – for textual documents; .tiff, .jpg – digital images; .mpg – digital videos; etc.)

Need to define "official NOAA document" vs. "Non-official NOAA document" (Official documents – those with author or office electronic signature authentication, official agency's banners, logos, etc.?) NOAA IR Copyright Unit (IRCU) team should determine a status of the Official NOAA Electronic Document?*
*The Digital Rights Management issue may also be addressed. (*Digital rights management (DRM) is, technically, the use of encryption (coding) of electronic data so that the creator has control over its use.)

**** "Digital signatures are a method of authenticating digital information analogous to ordinary physical signatures on paper, but implemented using techniques from the field of cryptography."
More on this in:
http://www.ncsconline.org/D_KIS/Trends/2005/DocManDigital.html

2. **Open ACCESS and INTEROPERABILITY.**
   The future NOAA Institutional Repository (NOAA IR) should be open to all NOAA employees and to the general public. It should be an open system*, interoperable with other related repositories and NOAALINC, the library online catalog. Therefore, the NOAA IR system should be able to support interoperability to provide access via multiple search engines and other discovery tools. It should provide no- or low-barrier access to NOAA intellectual products. However, some minimal restrictions on access may be considered; if adopted, these restrictions would define a multilevel scheme of access:
   - Temporary restrictions on access for particular scientific data (for example, due to a "grace period" for the Principal Investigators).
   - Internal only access for NOAA employees to some documents being in preliminary or draft stage, internal policy documents, etc. (for example, limiting access to departmental working papers to the members of that department)
   - Access limited to some NOAA officials only, especially for the NOAA-Military collaborative type of sensitive documents.

*Implementing these policy-based restrictions requires robust access and rights management mechanisms to allow or restrict access to the content by a variety of criteria, including user type, institutional affiliation, user community, and others.

* OCLC offers to its members free of charge the "Open Repository" framework software package:

1. **OpenURL v. 1.0** at: http://www.oclc.org/research/software/openurl/
(The open-source OpenURL 1.0 distributions consist of a jar file and a Java Servlet web application ("war" file) providing OpenURL 1.0 resolution capability. The default installation demonstrates its application by echoing OpenURL requests formatted in HTML. In practice, though, this service can be configured to support any context-sensitive service within the confines of the OpenURL 1.0 protocol).

2. **OAIHarvester2** at: ttp://www.oclc.org/research/software/oai/harvester2.htm (OAIHarvester2 Open Source Software (OSS) project is a Java application that provides an OAI-PMH harvester framework).

3. **OAICat** Open Source Software (OSS) at: http://www.oclc.org/research/software/oai/cat.htm (Java Servlet web application providing a repository framework that conforms to the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) v2.0. This framework can be customized to work with arbitrary data repositories by implementing some Java interfaces. Demonstration implementations of these interfaces are included in the webapp distribution).

3. COPYRIGHT.

All the NOAA documents being in public domain should be included in the NOAA IR database. Collaborative works should be cleared for copyright(s). To this end:

- A NOAA IR Copyright Unit (IRCU) should be established to assure clearing all copyright disputes.
- For all documents included in the NOAA IR, the IRCU team should ensure copyright compliance with the Copyright Act of 1976 and its 1998 Amendment.
- The NOAA Copyright Agreement Form should be developed to either transfer copyrights to NOAA, or to permit to publish collaborative work in the NOAA IR database. The AMS "Certification of Government Work Subject to Government Ownership" agreement my be used as a model. It is available online: http://ametsoc.org//pubs/copyrightinfo/AMS_copyright_crown.pdf
- A disclosure note on the "fair use" of the downloaded document should be included. (Erie's suggestion)
- All NOAA IR documents should comply with section 508 of the U.S. Rehabilitation Act of 1973 as amended in 1998 (29 U.S.C. 794d).
    - Include disclosure note for those PDF documents with graphic files (see Linda's e-mail)

**NOAA IRCU should be established and NESDIS General Counsel should be called for advice (?)*
NOAA Copyright Agreement Form should be automatically filed for every document entered into the IR.
Copyright aspect of contractors' publications as "work for hire" is stated in the Copyright Act of 1976: "If digital work created within scope of employment = work for hire. Institution owns [the work]. Statute says employer is the author."

4. **NOAA documents vs. NOAA funded documents**.
Both NOAA documents and NOAA-funded documents should be included in the NOAA IR database. However, the priority should be given to true NOAA official documents. For NOAA-funded documents, the copyright clearance procedure should be exercised.

5. OWNERSHIP.
The NOAA Central Library should own the NOAA IR.
All NOAA IR documents should be those in public domain (?) If so, then NOAA Central Library should maintain what public owns.

6. Withdrawals.

There may be unique circumstances when an IR document would be withdrawn. Documents may be withdrawn upon request of the author(s) or the appropriate NOAA office. An appropriate justification likewise should be made by the withdrawal-requesting person or office. An appropriate policy, which may reflect a process similar to weeding the library collection, should be written to specify when the document may or should be withdrawn from the NOAA IR database:

- Superseded documents should be marked or withdrawn.
- Scientifically incorrect ("bad science"), copyright infringement, allegations of plagiarism, etc. should be withdrawn.
- Sensitive documents may be withdrawn upon a request of the appropriate authorities (GPO, DHS, CIA, etc.)

**However, the trace of the document in the form of its metadata will be retained in the IR forever (?), along with remarks on why document is no longer available online.*

7. MANDATORY Status for NOAA electronic publishing.

The NOAA Institutional Repository should be a prestigious place to publish all NOAA official electronic documents. However, deposit herein also should be a mandatory procedure for all NOAA online publications regulated and implemented by the updated version of NOAA Administrative Order (NAO) 205-17 [and, perhaps, NAO201-32G?], approved and endorsed by appropriate high-level NOAA officials (viz., the NOAA Administrator and Directors of the NOAA line offices).

The NOAA IR, by capturing, preserving and disseminating the Agency's collective intellectual knowledge, serves as an important indicator of the institution's scientific [output and?] quality.

8. Benefits in publishing documents in NOAA IR for NOAA and the General Public:

- Provision of a centralized place to find all NOAA online documents, for NOAA scientists and for the general public.
- Assured long-term archiving, preservation, migration, etc.
- Sustainable online access to NOAA's intellectual capital.
- As a component of the global system of distributed, interoperable, related repositories for scholarly publishing, it offers the NOAA scientists greater visibilty.
- Includes the Publication Citing Status (PCS).
- Serves as an important tool in complementing NOAA's existing metrics to easily define institutional productivity and prestige. This can also translate into scientific quality of NOAA scholarship.
- No paper copy required for submission to the NOAA Central Library (save the library space, maintenance, processing labor, etc.)

Metadata issues:

1. Three types of metadata schemas considered:

A. **Administrative metadata** (context management-related)
[Administrative Metadata - Information that supports the management of a resource. The information libraries keep about acquisition, access restrictions, provenance, preservation and treatment decisions. For digital materials, administrative metadata includes information to:

- Determine who is allowed to use the objects and under what conditions,
- Track who owns the objects, who pays for the storage of the objects, who has permission to alter or delete the objects, etc., and
- Migrate the resource from one technical format to another (preservation)]

B. **Technical (Structural) metadata** (about the object itself)
[Technical Metadata comprises information that describes the capture process and technical characteristics of the digital texts or images. This standard … is an essential component of any digitization initiative for short-term and long-term management purposes. Technical metadata may include information that supports navigation among the components of a digital object. For example, turning pages of a book, jumping to a particular chapter or page, or switching between images and corresponding text. Structural metadata may be used by a computer program to generate an interface to an object, as for example, providing a way to view information from related statistical files together in a graphical way.] [Source of quote, if quoted material.]

Example of Technical metadata for still images, NISO Z39.87-2002 (http://www.niso.org/standards/resources/Z39_87_trial_use.pdf)

*Administrative and Technical Metadata are stored in the digital repository system

C. **Descriptive metadata** (describing the content) MARC, Dublin Core, MPEG-7, etc.
[Sometimes called intellectual or access metadata; descriptive metadata supports discovery and identification of objects. Title, author or creator, publisher, date of creation, etc. are examples of descriptive metadata. See below in section 4. In the context of the digital repository, the object owner is responsible for making descriptive metadata available by an appropriate catalog system (that is, an online catalog, a web site, etc.). Descriptive metadata in not stored in the Digital Repository.]

D. Preservation metadata (preservation related)

2. Automatic harvesting process for NOAA documents online
Any document that reflects the NOAA domain in its URL (.noaa.gov/xxxxx) should be considered for discovery [??] as a NOAA electronic document in the online environment. Automatic metadata generation would be

derived from documents stored with the digital object online (automatic harvesting metadata process)

3. Auto-load of existing MARC metadata records for NOAA electronic documents from NOAALINC to IR database should be considered:

a. Batch-load existing metadata from NOAALINC
b. Auto-feed, both ways, between two systems should be considered, as well as the auto-update of both systems (for edits, for example).
c. Mapping between NOAALINC and NOAA IR elements for easy import/export process and interoperability between two systems should also be considered.

4. **Proposed metadata elements considered for mapping?** (MARC/Dublin Core/MARCXML?)

- author(s)
- title
- publisher
- date issued
- series (authorized/traced version)
- keywords (whole metadata + full text searching)
- object/record control nos. (Persistent Identifiers)
  - System generated record ID,
  - ISSN/ISBN
  - URI/URL
  - NTIS
  - GPO Item No., etc.
- Abstract?
- Language
- Subject/descriptors

System requirements/file characteristics/format, etc. (technical characteristics of the digital object) – Tech. metadata info embedded

5. Authority control* - controlled names, places, terms used.

**It is very important that the future NOAA IR system provide users with controlled vocabulary searching. The controlled vocabulary "pick list" for keyword and subject searching is a minimum requirement.*

We also agreed that the IR system should be open source/open URL/OASIS concept compatible for easy access to the field-related repositories (See above, section 2. Open Access and Interoperability)

*Additional meeting to discuss Metadata issues for the future system may be called if needed.

DSpace home page is very up-to-date on many aspects of IR, Including policy-making, metadata, etc.
http://dspace.org/index.html
Some answers to the above questions by DSpace
http://dspace.org/faqs/index.html#community:
http://dspace.org/introduction/intro-faculty.html
An interesting Demo of the free DSpace system at:
http://dspace.org/implement/submit-content.html

* The readings Linda Pikula suggested are on the library's CDServer under:
\\Cdserver\Proposals\Institutional_Repository

Rev. 04/03/2006